Towards Broader Adoption of Massive Online Experiments

Joshua K. Hartshorne (joshua.hartshorne@bc.edu)
Department of Psychology, 140 Commonwealth Ave
Chestnut Hill, MA 02467 USA

Jesse Storbeck (jesse.storbeck@gmail.com)
Department of Psychology, 140 Commonwealth Ave
Chestnut Hill, MA 02467 USA

Keywords: online experiments; massive online experiments; citizen science; methodology

Significance

Since the Workshop on Scaling Cognitive Science was held at CogSci 2020, the methodological landscape of the field has undergone a significant shift toward online experiments. The COVID-19 pandemic forced greater numbers of cognitive scientists to collect data over the internet and to explore online alternatives to methods which they had previously used only in the lab. For example, since this time, eye-tracking capabilities have been introduced to a number of software packages for online experiments (Bánki, de Eccher, Falschlehner, Hoehl, & Markova, 2022; Slim & Hartsuiker, 2022; Yang & Krajbich, 2021).

As a result of the increased interest in online studies, the average cognitive scientist has never been more educated on web technologies than today. At the same time, cognitive scientists underutilize the affordances of online studies, still mostly running studies based on methods designed for the laboratory.

This full-day workshop explores the benefits, challenges, and implementation of *Massive Online Experiments* (MOEs), a methodology designed around the demands and opportunities of internet-based data collection. The goal of the workshop is to help more cognitive scientists begin to incorporate MOEs into their research.

What are MOEs, and Why Use them?

Massive Online Experiments use large samples of up to millions of subjects. Participants are recruited using a "citizen science" model, where subjects are part of the research team, studying themselves. They participate not for money but because the experiments are designed to be intrinsically motivating. This is not only more economical than paying millions of subjects, but empirically the number of people that will participate in a compelling science project is orders of magnitude more than the number that can be paid to do a boring online task.

The large, demographically diverse set of participants in MOEs allows researchers to design experiments that are not feasible in the traditional laboratory or online labor market model. MOEs enable statistically powerful comparisons across demographic variables like age, culture, and education level. Previous MOEs have simultaneously probed cross-cultural differences in domains such as moral reasoning, personality, and aesthetic preferences (Bleidorn et al., 2016; Gebauer et al., 2014; Reinecke & Gajos, 2014).

Moreover, large samples allow researchers to test the generalizability of findings across different theoretically-relevant parameters by running dozens of experimental variations simultaneously. When MOEs are combined with modern statistical analyses such as multi-level modeling, researchers can study generalization across large numbers of subjects and items simultaneously (Baribault et al., 2018).

The Challenges of Massive Online Experiments

There are two major challenges to broader adoption of MOEs. First, researchers are trained to think about experiments under the constraints of the traditional laboratory approach; however, the benefits of large-scale online studies are only realized in full if the experimental design is tailored to the format. For example, game-based approaches can be used to make tasks intrinsically rewarding for participants and can potentially produce higher quality data at lower cost (e.g., Long, Simson, Buxó-Lugo, Watson, & Mehr, 2023).

The second set of challenges are technical. Unlike labor market studies, there are no pre-existing platforms like MTurk for hosting MOEs. Researchers must create their own, along the lines of gameswithwords.org, labinthewild.org, or themusiclab.org. Studies may go viral, causing poorly-designed websites to crash. Assignment of items or conditions to subjects must be carefully done when there are tens of thousands of items, since random assignment at that volume will inevitably over- and under-sample some items. Displays and timing must be standardized across many devices. Etc.

Workshop Organization

The first half of this full-day workshop will include presentations on the benefits and challenges of MOEs and provide examples. The second half of the day will be a tutorial on software to implement MOEs: Pushkin.

Presenters span a range of CogSci topics and fields, including human-computer interaction, music cognition, language acquisition, decision-making, and social dynamics, illustrating the wide applicability of the workshop to the CogSci audience.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

Katharina Reinecke & Nigini Oliveira on Recruiting Volunteer Participants Online

Reinecke and Oliveira (associate professor and postdoc, respectively; Computer Science & Engineering, University of Washington) will report on more than 10 years of experience running the online experiment platform LabintheWild. Specifically, the talk will cover findings and lessons learned from their experiences conducting online studies with volunteers, including how participants benefit from these studies, how to ensure participants provide truthful answers, how to attract diverse samples, and how to estimate participants' viewing distance using Virtual Chinrest. They will conclude by showing how LabintheWild has enabled various research contributions (including in the fields of psychology, computer science, and medical science) that would have been difficult or impossible to achieve in laboratory studies or in studies conducted on online labor markets such as Mechanical Turk.

William Thompson on Experimental Evolution of a Cognitive Algorithm

Thompson (assistant professor, Psychology, UC-Berkeley) will discuss a large-scale behavioral experiment investigating social learning. Thousands of participants facing a sequential decision-making task were organized into generation-structured social networks and were allowed to learn from each other. He will discuss how knowledge evolved over time and talk about the practical challenges for such studies and how to address them with Dallinger.

Mathew Hardy on Biases in Social Networks

Hardy (Ph.D. student, Psychology, Princeton) will discuss a large behavioral experiment built using the Dallinger software package, which demonstrates that transmission through social networks can amplify perceptual biases. To address this, he draws on techniques from machine learning and statistics to identify a simple adjustment to transmission that is predicted to mitigate this amplification. In a second large experiment, this strategy reduced bias amplification while maintaining the benefits of information sharing.

Samuel Mehr on how Games Can Make Cognitive Science Better

Mehr (senior lecturer, Psychology, University of Auckland, and senior scientist, Haskins Laboratories) will talk about how citizen-science methods have demonstrated the potential for improving reproducibility and generalizability by recruiting large and diverse samples of participants and expanding stimulus sets and experiment types. He will discuss a key component of successful approaches to citizen science: the use of gamification techniques to bolster recruitment, increase the depth of participant engagement, and improve the ecological validity of data collected in experiments. Time will also be included for discussion of issues surrounding challenges gamifying research.

Joshua Hartshorne & Jesse Storbeck on Massive Online Experiments with Pushkin

For the second half of the workshop, Hartshorne and Storbeck (assistant professor and postdoc, respectively; Psychology & Neuroscience, Boston College) will present a tutorial on the implementation of a selection of topics covered in the preceding talks. Specifically, they will discuss Pushkin, a customizable, scalable ecosystem for MOEs. Pushkin address the additional challenges of MOEs relative to traditional online experiments, such as handling massive traffic fluctuations and efficiently distributing large numbers of items across large numbers of subjects. Pushkin provides many additional resources for maintaining an internet laboratory: a website template, an interactive forum, social media integration, and the ability for subjects to create an account (useful for longitudinal studies, etc.).

Acknowledgments

Funding provided by NSF 2229631 to JKH.

References

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam-and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 6162.

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., Van Ravenzwaaij, D., ... Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, *115*(11), 2607–2612.

Bleidorn, W., Schönbrodt, F., Gebauer, J. E., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2016). To live among like-minded others: Exploring the links between person-city personality fit and self-esteem. *Psychological science*, *27*(3), 419–427.

Gebauer, J. E., Bleidorn, W., Gosling, S. D., Rentfrow, P. J., Lamb, M. E., & Potter, J. (2014). Cross-cultural variations in big five relationships with religiosity: A sociocultural motives perspective. *Journal of Personality and Social Psychology*, *107*(6), 1064.

Long, B., Simson, J., Buxó-Lugo, A., Watson, D. G., & Mehr, S. A. (2023). How games can make behavioural science better. *Nature*, *613*(7944), 433–436.

Reinecke, K., & Gajos, K. Z. (2014). Quantifying visual preferences around the world. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 11–20).

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? a web-based replication of dijkgraaf, hartsuiker, and duyck (2017) using pcibex and webgazer. js. *Behavior Research Methods*, 1–19.

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505.